© 1998 Oxford University Press
Nucleic Acids Research, 1998, Vol. 26, No. 21 **4975–4982**

Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates

Joseph G. Hacia, Stephen A. Woski[1], Jacqueline Fidanza[2], Keith Edgemon, Nathaniel Hunt, Glenn McGall[2], Stephen P. A. Fodor[2] and Francis S. Collins*

National Human Genome Research Institute, National Institutes of Health, Building 31, Room 4B09, 31 Center Drive, Bethesda, MD 20892-2152, USA, [1]Department of Chemistry, University of Alabama, Tuscaloosa, AL 35487, USA and [2]Affymetrix, Santa Clara, CA 95051, USA

Received May 29, 1998; Revised and Accepted September 10, 1998

ABSTRACT

Pairs of high density oligonucleotide arrays (DNA chips) consisting of >96 000 oligonucleotides were designed to screen the entire 5.53 kb coding region of the hereditary breast and ovarian cancer *BRCA1* gene for all possible sequence changes in the homozygous and heterozygous states. Single-stranded RNA targets were generated by PCR amplification of individual *BRCA1* exons using primers containing T3 and T7 RNA polymerase promoter tails followed by *in vitro* transcription and partial fragmentation reactions. Fluorescent hybridization signals from targets containing the four natural bases to >5592 different fully complementary 25mer oligonucleotide probes on the chip varied over two orders of magnitude. To examine the thermodynamic contribution of rU·dA and rA·dT target·probe base pairs to this variability, modified uridine [5-methyluridine and 5-(1-propynyl)-uridine)] and modified adenosine (2,6-diaminopurine riboside) 5′-triphosphates were incorporated into *BRCA1* targets. Hybridization specificity was assessed based upon hybridization signals from >33 200 probes containing centrally localized single base pair mismatches relative to target sequence. Targets containing 5-methyluridine displayed promising localized enhancements in hybridization signal, especially in pyrimidine-rich target tracts, while maintaining single nucleotide mismatch hybridization specificities comparable with those of unmodified targets.

INTRODUCTION

Light-directed combinatorial chemical approaches allow the manufacture of high density arrays consisting of $>10^5$ distinct oligonucleotide species (20 μm feature size) on 1.2×1.2 cm^2 glass surfaces (1–2). Such arrays have been used to screen for mutations and polymorphisms in the *CFTR* gene (3), the HIV-1 reverse transcriptase and protease genes (4), the β-globin gene (5), the mitochondrial genome (6) and the *BRCA1* gene (7). Furthermore, they have been used to monitor gene expression (8), analyze gene function (9), optimize antisense oligonucleotide design (10) and acquire information from orthologous genes in related species (11).

A significant challenge in high density oligonucleotide array-based applications is to develop assay conditions so all fully complementary perfect match oligonucleotide probes of varying sequence content produce robust and specific target hybridization signals. Subsets of perfect match probes could have a greatly diminished hybridization signal due to decreased duplex stability resulting from sequence composition effects and inter- and intramolecular structures in both target and probes. When reliable data from such probes must be generated, hybridization conditions which are suboptimal for the specificity of other probes with robust hybridization signals may have to be employed. Herein, we analyze the affinity and specificity of RNA targets toward >90 000 oligonucleotide probes present on a pair of high density oligonucleotide arrays that scan the entire coding region of the *BRCA1* gene for all possible homozygous and heterozygous sequence changes. To pursue sequence composition effects on target hybridization and to explore possible solutions, we evaluated the effect of incorporating modified nucleoside triphosphates into *BRCA1* RNA target on hybridization signal and single nucleotide mismatch hybridization specificity.

MATERIALS AND METHODS

Synthesis of pyrimidine 5′-triphosphates

The synthesis of 5-(1-propynyl)-uridine was accomplished in three steps from commercially available 5-iodouridine (Sigma). First, treatment with excess acetic anhydride in pyridine produced tri-*O*-acetyl-5-iodouridine. This compound was converted to the 5-propynyl analog using the method of Hobbs (12). The free nucleoside was then generated by reaction with $NaOCH_3$ in methanol followed by desalting over BioRad AG-501 mixed bed resin.

The 5′-triphosphates of 5-(1-propynyl)-uridine and commercially available 5-methyluridine (R.I. Chemical) were synthesized using a two step procedure. First, conversions of the nucleosides

*To whom correspondence should be addressed. Tel: +1 301 496 0844; Fax: +1 301 402 0837; Email: fc23a@nih.gov

Table 1. *BRCA1* exon amplification primers

Exon	T3 Forward Primer Sequence (5′-3′)	T7 Reverse Primer Sequence (5′-3′)
2	d(ATTAACCCTCACTAAAGGGAAGTTGTCATTTTATAAACCTTT)	d(TAATACGACTCACTATAGGGATGTCTTTTCTTCCCTAGTATGT)
3	d(ATTAACCCTCACTAAAGGGAACGAACTTGAGGCCTTATG)	d(TAATACGACTCACTATAGGGATTGGATTTTCGTTCTCACTTA)
5	d(ATTAACCCTCACTAAAGGGACTCTTAAGGGCAGTTGTGAG)	d(TAATACGACTCACTATAGGGATTCCTACTGTGGTTGCTTCC)
6	d(ATTAACCCTCACTAAAGGGACTTATTTTAGTGTCCTTAAAAGG)	d(TAATACGACTCACTATAGGGATTTCATGGACATCACTTGAGTG)
7	d(ATTAACCCTCACTAAAGGGACAACAAAGAGCATACATAGGG)	d(TAATACGACTCACTATAGGGAGGGCTAAGGCAGGAGGACTGCT)
8	d(ATTAACCCTCACTAAAGGGATGTTAGCTGACTGATGATGGT)	d(TAATACGACTCACTATAGGGATCCAGCAATTATTATTAAATAC)
9	d(ATTAACCCTCACTAAAGGGACCACAGTAGATGCTCAGTAAATA)	d(TAATACGACTCACTATAGGGATAGGAAAATACCAGCTTCATAGA)
10	d(ATTAACCCTCACTAAAGGGATGGTCAGCTTTCTGTAATCG)	d(TAATACGACTCACTATAGGGAGTATCTACCCACTCTCTTCTTCAG)
11	d(ATTAACCCTCACTAAAGGGAATTAAATGAAAGAGTATGAGC)	d(TAATACGACTCACTATAGGGAGTGCTCCCAAAAGCATAAA)
12	d(ATTAACCCTCACTAAAGGGAGTCCTGCCAATGAGAAGAAA)	d(TAATACGACTCACTATAGGGATGTCAGCAAACCTAAGAATGT)
13	d(ATTAACCCTCACTAAAGGGAATGGAAAGCTTCTCAAAGTA)	d(TAATACGACTCACTATAGGGATGTTGGAGCTAGGTCCTTAC)
14	d(ATTAACCCTCACTAAAGGGACTAACCTGAATTATCACTATCA)	d(TAATACGACTCACTATAGGGAGTGTATAAATGCCTGTATGCA)
15	d(ATTAACCCTCACTAAAGGGATGGCTGCCCAGGAAGTATG)	d(TAATACGACTCACTATAGGGAACCAGAATATCTTTATGTAGGA)
16	d(ATTAACCCTCACTAAAGGGAATTCTTAACAGAGACCAGAAC)	d(TAATACGACTCACTATAGGGAAAACTCTTTCCAGAATGTTGT)
17	d(ATTAACCCTCACTAAAGGGAGTGTAGAACGTGCAGGATTG)	d(TAATACGACTCACTATAGGGATCGCCTCATGTGGTTTTA)
18	d(ATTAACCCTCACTAAAGGGAGGCTCTTTAGCTTCTTAGGAC)	d(TAATACGACTCACTATAGGGAGACCATTTTCCCAGCATC)
19	d(ATTAACCCTCACTAAAGGGATCTCCGTGAAAAAGAGC)	d(TAATACGACTCACTATAGGGACATTGTTAAGGAAAGTGGTGC)
20	d(ATTAACCCTCACTAAAGGGATATGACGTGTCTGCTCCAC)	d(TAATACGACTCACTATAGGGAGGGAATCCAAATTACACAGC)
21	d(ATTAACCCTCACTAAAGGGAATCATCAGGTGGTGAACAGAAG)	d(TAATACGACTCACTATAGGGAGTGCTGGAACTCTGGGGTTCT)
22	d(ATTAACCCTCACTAAAGGGATCCCATTGAGAGGTCTTGCT)	d(TAATACGACTCACTATAGGGAGAAGACTTCTGAGGCTAC)
23	d(ATTAACCCTCACTAAAGGGACAGAGCAAGACCCTGTCTC)	d(TAATACGACTCACTATAGGGACATTTTAGCCATTCA)
24	d(ATTAACCCTCACTAAAGGGATGATTAGAGCCTAGTCCAGG)	d(TAATACGACTCACTATAGGGATAGCCAGAAGTCCTTTTCAGG)

into the crude 5′-phosphodichlorodates were accomplished using the procedure of Sowa and Ouchi (13). Without isolation, these compounds were directly converted into the triphosphates by the addition of tributylammonium pyrophosphate and tributylamine in DMF (14). The triphosphates were purified by anion-exchange chromatography eluting with a gradient of triethylammonium formate (pH 6.5). Extensive lyophilization and co-evaporation with water provided the desired triphosphates as their triethylammonium salts. Identities of both compounds were confirmed by ^{31}P NMR spectroscopy and negative ion FAB-MS and the purities were determined to be >95% by analytical anion-exchange HPLC.

Data for 5-MeUTP: ^{31}P NMR (202 MHz, D_2O) δ –10.40 (d, J = 19.7 Hz, P_γ), –11.22 (dd, J = 20.4, 2.5 Hz, P_α), –22.84 (unresolved dd, $J_{apparent}$ = 19.8 Hz, P_β); MS (negative ion FAB) m/z 497 (100%, $[M^{4-} + 3H^+]$), 479 (24%, $[M^{4-} + 3H^+ – H_2O]$), 331 (82%).

Data for ᵖUTP: ^{31}P NMR (202 MHz, D_2O) δ –9.24 (m, P_α), –10.69 (d, J = 19.3 Hz, P_γ), –22.27 (unresolved dd, $J_{apparent}$ = 19.7 Hz, P_β); MS (negative ion FAB) m/z 521 (100%, $[M^{4-} + 3H^+]$).

NMR spectra were obtained on a Brüker AM500 spectrometer. MS were obtained on a VG Autospec mass spectrometer.

Synthesis of diaminopurine 5′-triphosphate (rDTP)

Diaminopurine-5′-monophosphate was synthesized from diaminopurine riboside (Reliable BioPharmaceuticals) using the described procedure (15) and purified on DEAE–Sephadex employing a 0–0.5 M LiCl gradient to give an 80% yield. The title compound (rDTP) was then synthesized from the monophosphate (16) in 80% yield, purified by RP-HPLC in 50 mM TEAA using an acetonitrile gradient and characterized by ^{31}P NMR and MS.

Data for DTP: ^{31}P NMR (D_2O, relative to H_3PO_4) δ –22.4 (t), –10.8 (d), –8.6 (broad, m). LCMS = 521.0 (M-H).

RNA target preparation

In vitro transcription templates were generated by PCR amplification of all *BRCA1* coding exons from genomic DNA using intronic forward and reverse primer pairs containing T3 and T7 promoter sequences, respectively (Table 1). Exon 11 was amplified using the EXPAND™ Long Range PCR Kit (Boehringer Mannheim) (7). The remaining 21 coding exons were amplified using the Amplitaq Gold PCR Kit (Perkin Elmer). Approximately 5 ng of each exon (except exon 11) were pooled and subject to T3 and T7 RNA polymerase *in vitro* transcription reactions. Exon 11 templates (~50 ng) were transcribed in a separate reaction. *In vitro* transcription reactions were performed in 20 µl reaction volumes using T3 RNA polymerase transcription buffer (Promega), 0.7 mM of the appropriate nucleoside triphosphates, 10 mM DTT, 0.7 mM biotin-16-CTP (Enzo Diagnostics) and 10 U T3 or T7 RNA polymerase as indicated. A 10 µl volume of pooled *BRCA1* exons was diluted into a 8 µl solution of 100 mM $MgCl_2$ with exon 11 transcription products separately treated in a like manner. Reactions were incubated at 94°C for 15 and 45 min, respectively, to fragment targets into ~50–100 nt long pieces which are more accessible to hybridization (7). In theory this produces a relatively random distribution of fragmentation products, however, some phosphodiester internucleotide linkages may be more reactive to hydrolysis than others. This may influence target hybridization in some sequence contexts and should be taken into consideration when interpreting array hybridization data in this study.

Array hybridization and data collection

Fragmented exon 11 and pooled exon targets were combined and diluted into a 400 µl volume of hybridization buffer A (3 M tetramethylammonium chloride, 1× TE, pH 7.4, 0.001% Triton X-100) or B (6× SSPE, 0.005% Triton X-100) containing either 1 nM 5′-fluorescein-labeled control oligodeoxyribonucleotides S (5′-CGGTAGCATCTTGAC-3′) or AS (5′-GTCAAGATGC-TACCG-3′) (for arrays complementary to sense and antisense strand targets, respectively) (7). Arrays were hybridized with target, stained with a phycoerythrin–streptavidin conjugate (Molecular Probes) and hybridization signals quantitated as previously described (7).

RESULTS AND DISCUSSION

Oligonucleotide array design

Extending previous analysis of the 3.43 kb central *BRCA1* exon 11 (7), a pair of arrays consisting of >96 600 oligonucleotides was designed to scan both strands of the 5.59 kb *BRCA1* coding sequence (containing 22 coding exons) for all possible sequence

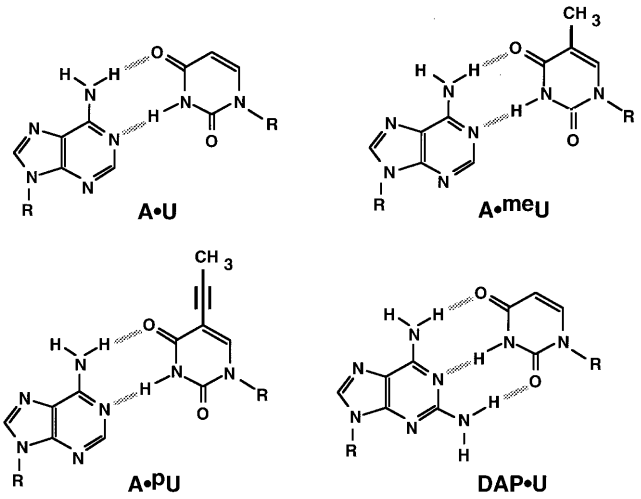

Figure 1. Proposed hydrogen bonding schemes of modified pyrimidine and purine base pairs. ^{me}U, ^{p}U and DAP represent 5-methyluridine, 5-(1-propynyl)-uridine and 2,6-diaminopurine, respectively. Dashed lines represent proposed hydrogen bonds.

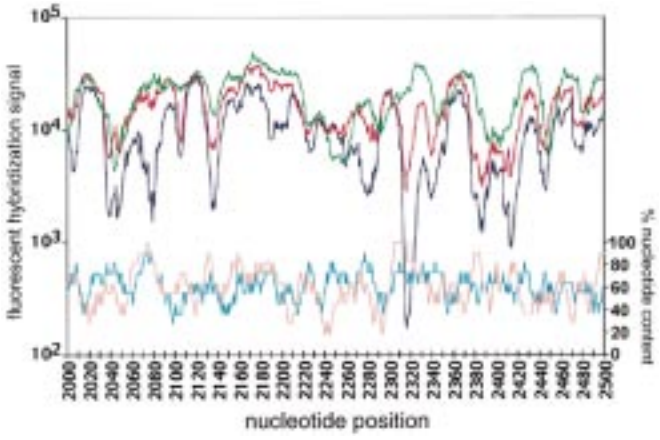

Figure 2. Hybridization signal intensities of modified pyrimidine antisense targets. Averaged fluorescence intensities (two experiments) of targets hybridized (buffer A, 40°C) to arrayed *BRCA1* perfect match probes are shown. Fluorescent signal intensities from perfect match probes complementary to nt 2500–3000 of the *BRCA1* cDNA sequence are plotted on the log scale *y*-axis with the corresponding nucleotide position listed on the *x*-axis. Dark blue, dark red and green lines represent data from unmodified, ^{me}U and ^{p}U targets respectively. Rolling average percentages (9 nt window size) of array A·T and A·G content are plotted in light blue and pink respectively relative to the right *y*-axis.

changes not involving insertions and deletions greater than the probe length. Every *BRCA1* nucleotide position is interrogated by four 25 nt probes on the chip, each substituted with one of the four nucleotides in the central position. The ratio of hybridization signal to perfect match probes relative to those of the three single nucleotide substitution mismatch probes provides a measurement of hybridization specificity under a given set of conditions (3,4,6,7). A set of perfect match oligonucleotides (25, 23 and 22 nt in length) per target strand base form a contig of single nucleotide overlapping probes which tile across all *BRCA1* coding exons along with 10 bp of flanking intronic sequences. These probes are used in two color hybridization mutational analysis experiments (6,7).

Modified nucleoside triphosphate design

To directly test the effects of target sequence composition due to dA·rU and dT·rA probe·target interactions on the range of hybridization signals, we incorporated modified nucleotides into *BRCA1* targets. Elegant studies have shown that several RNA polymerases can tolerate template-directed incorporation of non-natural nucleoside triphosphates (17,18). Furthermore, modified uridine derivatives have been characterized which enhance hybridization with adenosine-rich targets including 5-methyluridine (^{me}U) (19) and 5-(1-propynyl)-deoxyuridine (^{p}dU). The former is a naturally occurring post-transcriptional modification in several tRNA species (20) while the latter has been employed in antisense oligonucleotide gene expression inhibition studies (21). The enhanced thermodynamic stability of these modified uridine-containing base pairs, shown in Figure 1, has been postulated to be due to more favorable stacking interactions and entropic factors (21,22). These entropic factors may stem from the enhanced displacement of highly ordered water molecules from the duplex due to these modified uridines (21,22). 2,6-Diaminopurine (DAP) is a modified adenine base which enhances binding affinity to thymine although having significant affinity to other bases in some sequence contexts (23,24). This modified adenine has been proposed to increase the stability of base pairs with thymidine (Fig. 1) due to altered stacking interactions, the formation of an additional hydrogen bond in the modified base pair and removing the spine of hydration in the minor groove (23,24; Fig. 1).

Effects of modified nucleoside triphosphates on *in vitro* transcription reactions

In vitro transcription reactions were performed in the presence of ATP, CTP (including biotin-15-CTP for post-hybridization array staining with a phycoerythrin–streptavidin conjugate), GTP and UTP. When examining 5-methyl-UTP (^{me}UTP) and 5-(1-propynyl)-UTP (^{p}UTP) incorporation, UTP was excluded. Likewise, ATP was excluded when transcribing with diaminopurine riboside 5′-triphosphate (DTP). For T3 and T7 RNA polymerase *in vitro* transcription reactions, ^{me}UTP did not significantly effect transcription yield relative to UTP based on ethidium bromide staining of the 3.43 kb exon 11 transcription products on agarose gels (data not shown). Substitution of ^{p}UTP for UTP in T3 RNA polymerase-mediated transcription reactions caused an ~4-fold decrease in ethidium bromide stained reaction exon 11 transcription product. In the analogous T7 RNA polymerase-mediated reaction, an ~10-fold decrease in exon 11 transcription product was found. DTP caused a more dramatic decrease in exon 11 transcription product, 10- and 20-fold for T3- and T7-mediated transcription reactions, respectively. All reactions containing both DTP and ^{me}UTP or ^{p}UTP showed greatly diminished transcription product yields (>50-fold).

Equivolume amounts of transcription products were fragmented and diluted into hybridization buffer (Materials and Methods). Target concentrations were not adjusted since it is not cost practical to increase transcription reaction volumes for ^{p}UTP and DTP *in vitro* transcription reactions. Furthermore, co-transcription of the smaller exons increases sample throughput and decreases reagent usage but produces a mixture of RNA species which are difficult to resolve and quantitate by gel electrophoresis. In theory

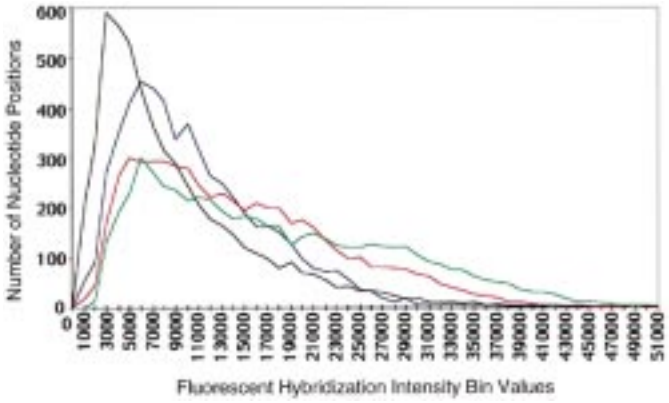

Figure 3. Binned hybridization signal intensities of modified antisense targets. Averaged fluorescent hybridization signal intensities of antisense targets (two experiments) to perfect match probes (buffer A, 40°C) are calculated and placed into bin values listed on the *x*-axis. The number of perfect match probes within these bin values (each encompassing 1000 fluorescence intensity units greater than listed value, except for the first two bins which correspond to perfect match probes having between 0 and 500 and those having between 500 and 1000 fluorescence intensity units) is shown on the *y*-axis. Dark blue, dark red, green and black lines represent data acquired from unmodified, ^{me}U, ^{p}U and DAP targets respectively.

it is possible to normalize the concentration of unmodified targets relative to the 5-(1-propynyl)-uridine and 2,6-diaminopurine modified targets. However, this would globally reduce hybridization for unmodified targets, making it difficult to accurately quantitate signals from all oligonucleotide probes. Therefore, the results generated from 5-(1-propynyl)-uridine and 2,6-diaminopurine (and to a lesser extent 5-methyluridine) containing targets should be taken as qualitative rather than quantitative. Our goal was to elucidate how modified nucleoside triphosphate incorporation would affect the performance of *BRCA1* target hybridization within the context of previously established assay conditions (7).

Hybridization properties of unmodified targets

The intensity of specific target hybridization may be shown by plotting hybridization signal strength to each perfect match probe per nucleotide position (Fig. 2). The fluorescent hybridization signal of perfect match probes varies over 130- and 230-fold (sense strand data) and 250- and 620-fold (antisense strand data) for unmodified target at 40°C in buffers A and B, respectively (based on averages of the 10 highest and lowest hybridization signals). Localized decreases in hybridization signal cannot be fully accounted for by thermodynamic parameters based upon target A/U or pyrimidine content (Fig. 2) but presumably also reflect potential intra- and intermolecular target and/or probe structures that inhibit hybridization (10). Of note is the use of tetramethylammonium (TMA) salts (25) which, along with betaine (26), have been widely used to minimize differences in oligonucleotide hybridization due to A·T content. These effects have been attributed to a non-cooperative differential stabilization of A·T (or A·U) base pairs, relative to G·C base pairs within duplex nucleic acid (25). Nevertheless, these buffers do not completely ameliorate energetic differences in the hybridization of short oligonucleotide targets (27). Although TMA^+ counterions altered hybridization to subsets of arrayed oligonucleotides relative to Na^+ counterions, they did not produce a globally uniform hybridization pattern (data not shown).

Table 2. Sequence tracts with enhanced hybridization due to analog incorporation

Sub[1]	Strand[2]	Tract[3]	Sequence[4]	Gain[5]
^{me}U	S	3062-3091	ACCACTTTTTCCCATCAAGTTCATTTTGTT	7.2
	S	3488-3509	cagaTTTCTCTCCATATCTGATTTCAgata	6.5
	S	3799-3808	ccctgcttccAACACTTGTTatttggtaaa	5.6
	S	2984-2988	ggttttgtctatcATCTCagttcagaggca	4.8
	S	1652-1665	acatcaggCCTTCATCCTGAGGattttatc	4.5
^{p}U	S	380- 408	tGAAATCATTTGTGCTTTTCAGCTTGACAc	8.8
	S	3491-3508	agatttCTCTCCATATCTGATTTCagataa	4.3
	S	355- 364	agtacgagatTTAGTCAACTtgttgaagag	3.3
	S	4828-4857	TCTTCTCTGATGACCCTGAATCTGATCCTT	3.2
	S	5531-5543	cacaggtgtCCACCCAATTGTGgttgtgca	2.9
DAP	S	826- 849	ataCTGAACATCATCAACCCAGTAATAatg	7.3
	S	3213-3220	taataacattAGAGAAAAtgtttttaaaga	6.9
	S	997-1003	gtttattactcaCTAAAGAcagaatgaatg	6.2
	S	1618-1623	gtcccctcacaaATAAATtaaagcgtaaaa	5.9
	S	1047-1053	ataCTGAACATCATCAACCCAGTAATAatg	5.6
^{me}U	AS	2309-2322	agagaagaAAAAGAAGAGAAACtagaaaca	20.1
	AS	4154-4164	cagatgatgaAGAAAGAGGAAcgggcttga	11.3
	AS	3448-3460	aaataaaaaAGCAAGAATATGAagaagtag	10.9
	AS	1626-1632	acaaaataaattaAAGCGTAaaaggagacct	10.9
	AS	2071-2076	gtgaagagataaAGAAAAaaaagtacaaca	9.7
^{p}U	AS	2309-2321	agagaagaAAAAGAAGAGAAActagaaaca	174.0
	AS	3160-3165	ctgaaagagaaaTGGGAAatgagaacattc	35.0
	AS	3216-3224	ataacattagaGAAAATGTTtttaaagaag	30.5
	AS	4155-4161	tcagatgatgaaGAAAGAGgaacgggcttg	28.6
	AS	3450-3458	aaataaaaaagCAAGAATATgaagaagtag	27.7
DAP	AS	3634-3655	taaggAAAGTTCTGCTGTTTTTAGCAAaag	5.0
	AS	3893-3912	gaggaGAATTTATTATCATTGAAGAatagc	4.6
	AS	254- 263	*TAATTTATAG*TTTTGCATGctgaaactt	4.3
	AS	3204-3233	AATAACATTAGAGAAAATGTTTTTAAAGAA	4.1
	AS	1433-1450	tcatgaGGCTTTAATATGTAAAAGtgaaag	4.0

[1]Target substitutions with ^{me}U, 5-methyluridine and ^{p}U, 5-(1-propynyl)-uridine.
[2]S, sense and AS, antisense strand data.
[3]Exonic nucleotide tracts with peak hybridization enhancements relative to unmodified targets (40°C, Buffer A).
[4]Tracts (*BRCA1* sense strand nucleotide sequence shown) displaying the five largest hybridization signal increases relative to unmodified targets. Italicized letters indicate intronic sequence and uppercase letters represent nucleotides with highest enhancement levels.
[5]Maximum level of hybridization signal enhancement.

Hybridization properties of modified targets

The range of fluorescent signal intensities was narrowed somewhat by ^{me}U and ^{p}U incorporation (Fig. 2). Global effects on the hybridization signal strength of ^{me}U, ^{p}U and DAP modified targets were further characterized by calculating the number of nucleotide positions having signal intensities within categorized bins (Fig. 3). Relative to unmodified antisense target, ^{me}U- and ^{p}U-substituted targets showed hybridization signals shifted towards overall higher values while those containing DAP shifted towards lower values. Relative to unmodified sense strand target, ^{me}U- and DAP-substituted targets showed similar average hybridization signal intensities (differing <1.2-fold), however, ^{p}U-substituted target had an ~2-fold decreased hybridization signal (buffer A, 40°C). This may result from lowered transcription reaction yields as well as from intra- and/or intermolecular target structures.

To assess localized changes in hybridization signal due to modified nucleotide incorporation, we calculated the ratio of perfect match probe hybridization signal intensity of modified to

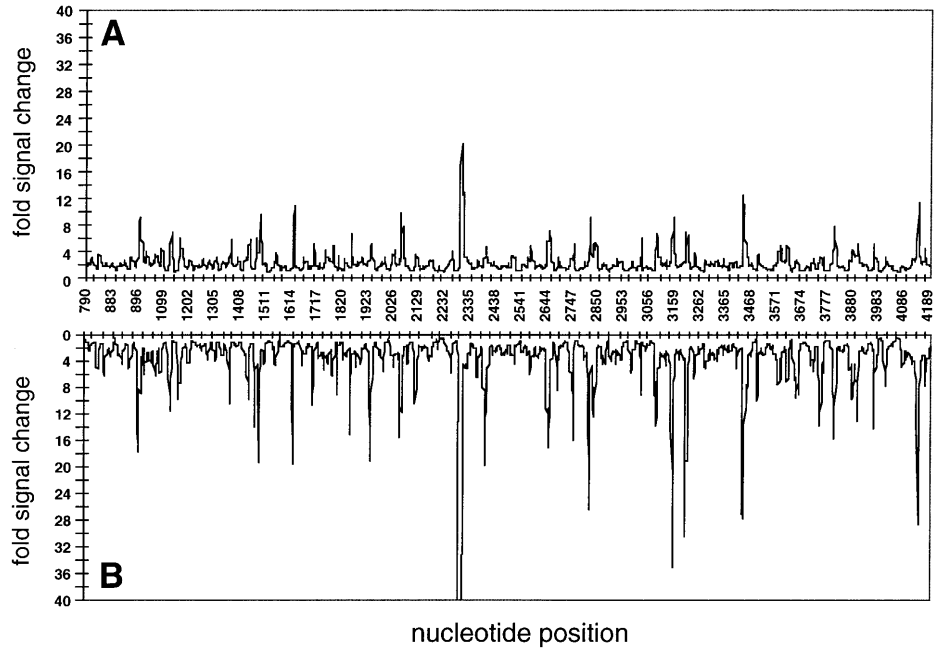

Figure 4. Relative hybridization intensities of modified pyrimidine antisense targets. Average fluorescence intensities of targets (two experiments) hybridized (buffer A, 40°C) to perfect match probes corresponding to exon 11 of the *BRCA1* oligonucleotide arrays were quantitated. Ratio of perfect match probe hybridization signals of (**A**) ^{me}U target relative to unmodified target and (**B**) ^{P}U target relative to unmodified target are given.

unmodified target for each sense and antisense strand nucleotide (Fig. 4). The largest signal enhancements in ^{me}U- and ^{P}U-substituted targets were primarily found in pyrimidine-enriched tracts containing a large number of uridine residues (Table 2). Homopyrimidine tracts are preferentially localized on the antisense *BRCA1* strand with 15 sequence tracts ≥10 nt long and only one such homopurine tract on this strand. Since DNA·RNA hybrid duplexes containing homopyrimidine RNA tracts are less stable than hybrid duplexes of identical sequence containing homopurine RNA tracts (28), modified uridine analogs have the best opportunity to significantly affect hybridization in these sequence contexts (Fig. 2). This could explain why ^{me}U and ^{P}U substitutions have a greater positive impact on hybridization signal from antisense than from sense target strands. While DAP-substituted targets also show regions of enhanced signal (Table 2) they are significantly less pronounced than the highest found in the pyrimidine-substituted targets. This may reflect the relative stability of DNA·RNA hybrids containing unmodified homopurine RNA strands.

Unmodified and ^{me}U-substituted targets show similar single nucleotide mismatch destabilization properties on both strands under a variety of assay conditions (Table 3 and Figs 5 and 6). Significantly, this selectivity is maintained with the antisense strand where ^{me}U incorporation produced the greatest localized hybridization signal enhancements (Fig. 4 and Table 2). When comparing hybridization properties of ^{me}U-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was only a 2.3-fold average decrease in single nucleotide mismatch discrimination in the five antisense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement. Lower temperatures are used in this comparison for unmodified targets due to diminished, thus less reliable, hybridization signal in these sequence tracts at 42°C.

The global single nucleotide mismatch specificity of probe·target interactions decreases when comparing ^{P}U-substituted (42°C) and unmodified targets under identical hybridization conditions (Table 3 and Fig. 5). This decreased specificity is highlighted in areas of the greatest probability of ^{P}U incorporation. When comparing hybridization properties of ^{P}U-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was a 5.1-fold average decrease in single nucleotide mismatch discrimination in the five antisense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement.

An example of the localized effects of modified pyrimidine substitutions on target hybridization signal strength from perfect match and single nucleotide mismatch probes is shown in Figure 6. While ^{me}U incorporation enhances hybridization signal and maintains a specific hybridization pattern to match probes in both sequence contexts, ^{P}U incorporation increases hybridization signal at the expense of single nucleotide mismatch specificity (Fig. 6c and f). Importantly, significantly increased ^{P}U- substituted target cross-hybridization often occurs in all four probes interrogating a single target nucleotide. Increased cross-hybridization to other areas of the array is also found with ^{P}U-substituted relative to unmodified oligodeoxyribonucleotides when hybridized in buffer A (data not shown).

DAP-substituted sense targets produce significantly stronger cross-hybridization to single base pair mismatch probes than unmodified targets (Table 3). Nevertheless, when comparing hybridization properties of DAP-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was a only

Table 3. Single nucleotide mismatch specificity ratios

		unmodified				5-methyluridine				5-(1-propynyl)-uridine				2,6-diaminopurine		
	Temp °C[1]	35	40	42	45	35	40	42	45	35	40	42	45	35	40	42
	Bin[2]	% Frequency[3]				% Frequency				% Frequency				% Frequency		
Buffer A Sense Data	< 1.2	1.4	0.3	0.4	0.3	1.5	0.5	0.4	0.4	7.2	3.5	2.6	1.5	20.4	13.9	17.8
	1.2 - 2.0	25.5	16.0	13.8	6.5	32.0	22.9	21.3	18.3	35.2	29.1	21.4	16.5	43.3	45.0	34.9
	2.0 - 3.0	30.7	30.3	32.0	24.8	36.4	39.2	37.6	41.5	33.2	39.0	41.7	43.9	20.9	22.8	25.4
	3.0 - 4.0	17.6	20.5	23.3	25.5	16.3	19.8	21.0	22.7	14.3	16.8	22.0	24.1	8.0	8.8	11.1
	4.0 - 5.0	10.5	13.1	13.5	17.5	7.0	8.8	10.4	10.1	5.5	6.6	7.6	8.7	3.8	4.6	5.3
	5.0 - 6.0	5.7	7.6	7.3	10.6	3.2	4.3	4.5	3.8	2.4	2.6	2.7	2.8	1.6	2.3	2.8
	6.0 - 7.0	3.4	4.2	4.0	6.1	1.5	2.1	2.2	1.6	1.1	1.2	1.3	1.3	0.8	1.2	1.3
	7.0 - 8.0	2.1	3.1	2.1	3.3	0.9	1.0	1.2	0.8	0.5	0.5	0.5	0.7	0.4	0.6	0.6
	8.0 - 9.0	1.0	1.7	1.3	2.3	0.5	0.6	0.6	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4
	9.0 - 10.0	0.7	0.9	0.6	1.1	0.2	0.3	0.4	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.2
	> 10.0	1.5	2.3	1.7	2.2	0.5	0.7	0.5	0.3	0.2	0.3	0.1	0.1	0.4	0.3	0.3
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer B Sense Data	< 1.2	3.6	1.1	1.0	0.9	3.2	1.3	0.6	0.5	6.4	1.6	1.2	1.1	46.4	22.3	19.2
	1.2 - 2.0	38.8	26.1	24.7	24.7	38.6	34.5	19.3	31.0	39.3	26.5	26.1	20.9	39.2	49.1	50.1
	2.0 - 3.0	28.0	31.9	32.1	32.2	30.6	34.3	34.6	37.9	32.3	38.4	37.6	41.4	10.3	18.9	19.6
	3.0 - 4.0	13.5	18.8	19.1	19.3	14.3	15.8	21.6	16.9	13.0	18.8	18.6	20.9	2.7	6.2	6.5
	4.0 - 5.0	7.1	10.7	11.0	10.7	6.3	7.3	12.1	7.2	5.3	8.0	8.3	9.1	0.8	1.9	2.6
	5.0 - 6.0	3.4	5.4	5.8	5.6	3.3	3.2	5.6	3.2	2.1	3.6	4.2	3.9	0.5	0.9	1.0
	6.0 - 7.0	2.0	2.8	2.8	2.9	1.6	1.8	2.8	1.5	0.9	1.6	1.9	1.4	0.1	0.4	0.6
	7.0 - 8.0	1.3	1.5	1.5	1.5	1.0	0.8	1.5	0.8	0.4	0.7	0.9	0.8	0.1	0.2	0.2
	8.0 - 9.0	0.7	0.7	0.8	1.0	0.6	0.5	0.8	0.5	0.2	0.4	0.5	0.3	0.0	0.1	0.1
	9.0 - 10.0	0.4	0.4	0.5	0.6	0.3	0.3	0.5	0.2	0.1	0.2	0.3	0.1	0.0	0.0	0.1
	> 10.0	1.1	0.7	0.6	0.7	0.3	0.3	0.8	0.2	0.1	0.2	0.3	0.1	0.0	0.0	0.1
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer A Antisense Data	< 1.2	0.8	0.2	0.3	0.5	1.1	0.1	0.2	0.3	9.3	3.8	3.1	1.1	2.9	0.9	0.5
	1.2 - 2.0	16.6	7.4	4.3	4.0	17.6	11.5	7.5	7.1	40.5	32.0	27.5	16.8	24.0	14.8	15.1
	2.0 - 3.0	28.8	20.8	14.9	17.2	33.1	31.2	28.0	29.7	34.4	37.1	39.0	39.1	30.6	26.3	33.2
	3.0 - 4.0	19.2	21.1	20.9	26.6	23.2	24.2	24.1	28.5	10.8	16.0	18.4	24.7	18.7	20.7	22.2
	4.0 - 5.0	12.1	15.6	17.2	20.8	11.8	13.4	16.4	17.0	3.2	6.1	7.0	11.0	10.0	13.5	11.7
	5.0 - 6.0	7.4	10.4	12.8	12.5	6.0	7.5	8.8	8.7	1.0	2.6	2.7	4.3	5.1	8.3	6.7
	6.0 - 7.0	4.2	7.3	8.7	7.2	3.0	4.7	5.7	4.4	0.5	1.1	1.2	1.7	3.0	5.4	4.0
	7.0 - 8.0	3.3	5.2	6.0	4.2	1.4	2.7	3.3	2.2	0.2	0.6	0.5	0.8	1.8	3.5	2.2
	8.0 - 9.0	2.0	3.1	4.1	2.6	1.0	1.6	2.2	1.2	0.1	0.2	0.2	0.3	1.1	2.2	1.2
	9.0 - 10.0	1.4	2.4	2.9	1.7	0.6	1.0	1.2	0.4	0.0	0.2	0.1	0.1	0.7	1.3	0.9
	> 10.0	4.2	6.7	7.9	2.8	1.1	2.0	2.5	0.5	0.1	0.3	0.2	0.2	2.1	3.2	2.2
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer B Antisense Data	< 1.2	1.7	1.6	1.8	2.6	0.9	0.3	0.2	0.4	3.2	0.7	0.5	0.5	7.2	3.9	2.8
	1.2 - 2.0	14.1	13.3	11.2	15.2	16.3	13.0	12.7	12.8	35.1	17.4	17.5	21.4	32.5	26.7	24.2
	2.0 - 3.0	26.2	25.6	23.6	29.8	31.4	31.8	33.0	31.5	37.7	36.0	36.7	41.4	28.0	29.8	30.4
	3.0 - 4.0	19.4	19.6	20.2	22.5	21.7	22.5	22.6	23.6	14.1	21.9	20.9	19.4	14.0	16.1	17.2
	4.0 - 5.0	12.5	12.9	13.9	13.2	11.6	13.3	13.5	14.4	5.7	10.9	11.1	9.1	7.6	8.4	9.6
	5.0 - 6.0	8.1	8.6	9.2	7.4	6.9	7.6	7.2	8.4	2.1	5.3	5.5	4.5	3.6	4.9	5.2
	6.0 - 7.0	5.4	5.6	6.0	4.2	4.0	4.5	4.5	3.8	0.9	3.0	3.3	1.9	2.3	3.0	3.4
	7.0 - 8.0	3.5	3.6	4.1	2.3	2.5	2.6	2.4	2.1	0.5	1.9	1.9	0.9	1.5	2.2	2.2
	8.0 - 9.0	2.6	2.5	3.0	1.3	1.7	1.7	1.6	1.2	0.2	1.1	1.1	0.5	1.0	1.3	1.6
	9.0 - 10.0	1.8	1.9	1.9	0.6	1.0	1.0	0.7	0.8	0.1	0.5	0.6	0.2	0.7	0.9	1.0
	> 10.0	4.8	4.8	5.0	0.8	1.9	1.7	1.5	1.1	0.3	1.1	0.9	0.2	1.7	2.8	2.3

[1]Hybridization temperature.
[2]Single nucleotide mismatch specificity ratio bins.
[3]% of BRCA1 coding nucleotide positions within the specified ratio bin.

a 2.3-fold average decrease in single nucleotide mismatch discrimination in the five sense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement. Therefore, cross-hybridization occurs in different sequence contexts distributed throughout the array. Furthermore, DAP-substituted antisense targets showed single nucleotide mismatch specificities similar to unmodified targets, presumably due to the decreased number of adenine-rich tracts on this strand and thus lower levels of DAP incorporation. For both sense and antisense analysis, lower overall transcriptional yields of DAP-modified targets result in a lower target concentration in the hybridization reaction. This increases the stringency of the hybridization reaction and consequently increases single nucleotide mismatch discrimination. Increasing the DAP-modified target concentration to produce more robust hybridization signals, especially for antisense strand targets, will result in lower single nucleotide discrimination, thus reducing the usefulness of this modification.

Co-incorporation of DAP and modified uridine residues into sense and antisense targets had a significantly negative effect on hybridization specificity (data not shown). A large component of this problem is the decreased transcription product yield (>50-fold). It does not appear feasible to simultaneously incorporate modified pyrimidine and purine nucleotide triphosphates and retain robust product yield which has non-degenerate hybridization specificity with these enzymes under these conditions. In the future, RNA polymerases may be engineered to have increased ability to

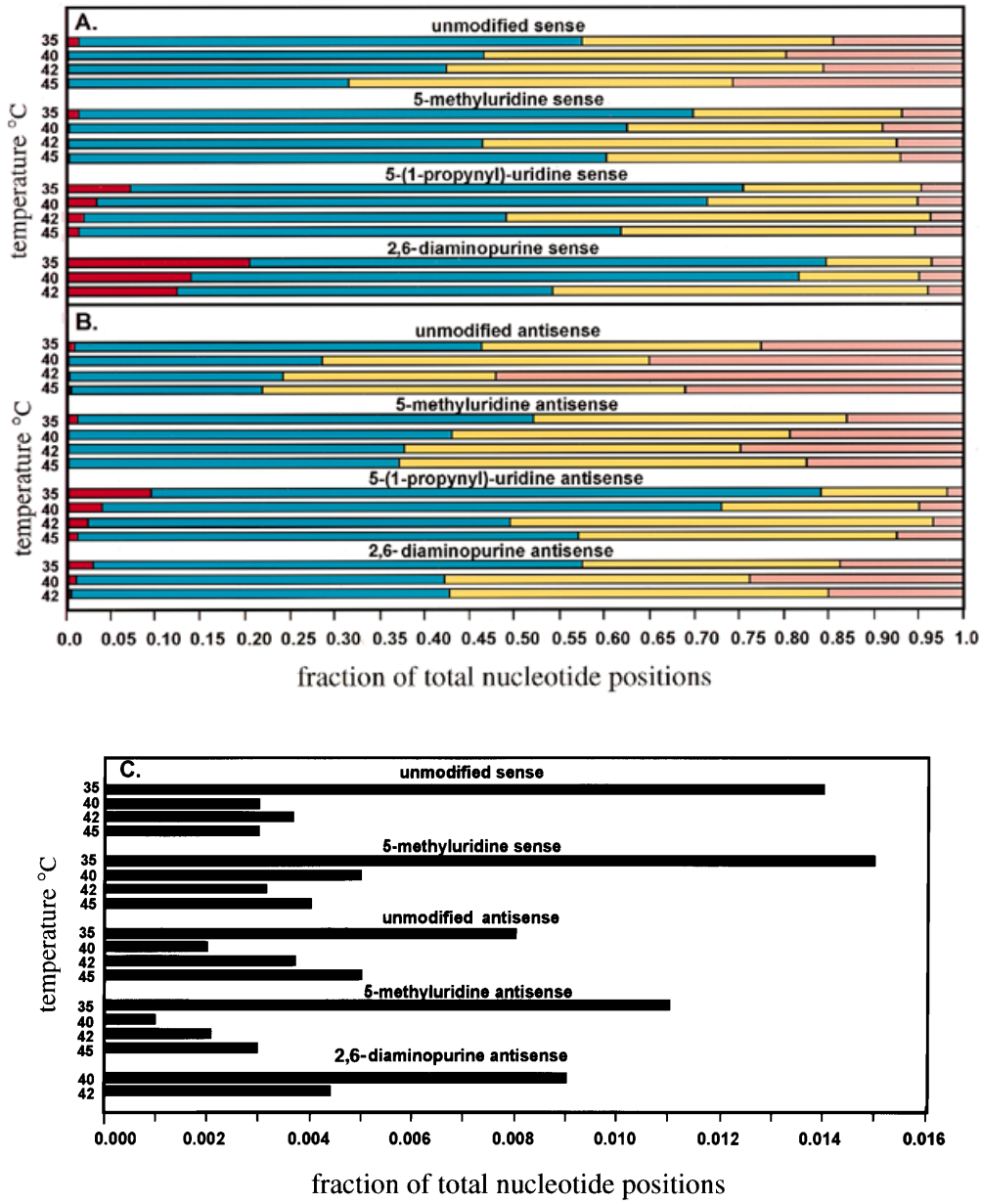

Figure 5. Single nucleotide mismatch specificity ratios of modified targets. Averaged ratios of target hybridization signals from perfect match probe to next highest mismatch probe taken from a minimum of two experiments (buffer A, 40°C). The *x*-axis represents the fraction of nucleotide positions within categorized bins of single nucleotide mismatch ratios <1.2 (red bars), between 1.2 and 3 (blue bars), between 3 and 5 (yellow bars) and >5 (orange bars), respectively. Sense and antisense strand data are shown in (**A**) and (**B**), respectively. Selected data of the single nucleotide mismatch ratios <1.2 are shown on a magnified *x*-axis in (**C**).

incorporate modified nucleoside triphosphates into transcription products.

Potential applications for modified targets

The examined modified triphosphates can be incorporated into large RNA transcripts by T3 and T7 RNA polymerases. While DAP and ^{P}U incorporation both lead to enhanced hybridization in specific sequence contexts, the loss of binding specificity reduces the likelihood of their use in mutation screening analysis. ^{me}U incorporation enhances target hybridization signals within specific sequence contexts and does not substantially increase hybridization signal to single nucleotide mismatch probes. Enhanced ^{me}U RNA target hybridization signals will be especially important in the mutational analysis of genes having localized regions of strongly biased A·T sequence content (i.e. *BRCA2* and *ATM* genes). Furthermore, in conjunction with modified oligonucleotide surface probes (Fidanza *et al*., unpublished observations) it may be possible to normalize the binding affinity of all perfect match probes in an array (29). This would potentially allow hybridization conditions to be used which globally optimize hybridization signal strength and specificity.

Modified nucleoside triphosphate usage could also benefit RNA expression monitoring experiments based on hybridization

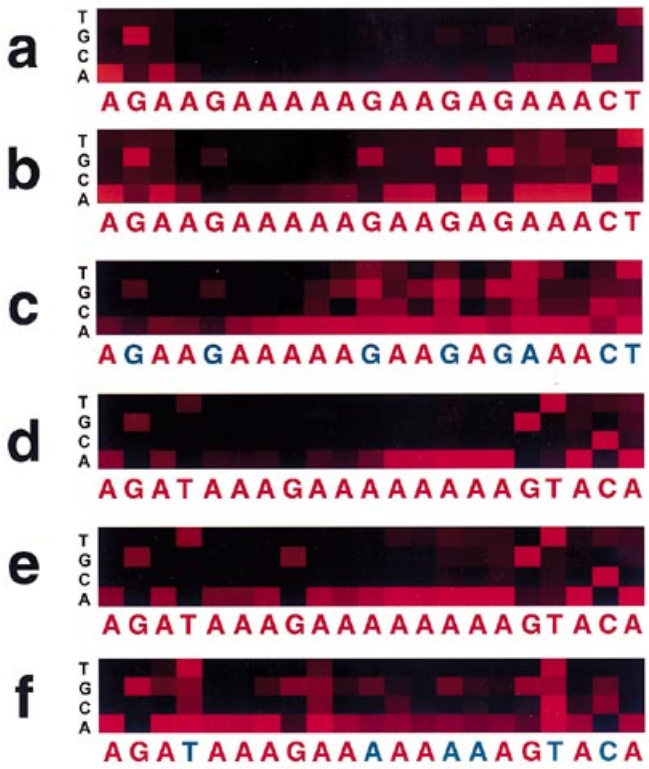

Figure 6. Modified pyrimidine target image comparisons. Magnified digitized false color images showing hybridization pattern of *BRCA1* antisense targets (buffer A, 40°C). Brightness and contrast settings are changed in each panel to increase image clarity. Nucleotide identities, determined through dideoxy-sequencing analysis, are given under the respective column and colored red or blue if correctly or incorrectly identified by hybridization analysis (perfect match probe intensity being at least 1.2-fold greater than that of the next highest mismatch probe intensity), respectively. Several base calls may be difficult to visualize due to limitations in printing technology as well as in the linear range of the human eye for detecting monochromatic color changes. (**a**–**c**) Hybridization patterns of unmodified, ^{me}U and ^{p}U antisense strand targets, respectively, to nt 2303–2323 of *BRCA1* cDNA. (**d**–**f**) Hybridization patterns of unmodified, ^{me}U and ^{p}U antisense strand targets, respectively, to nt 2065–2085 of *BRCA1* cDNA.

to high density oligonucleotide arrays (8). In such experiments, perfect match probe oligonucleotides are selected based upon sequence composition effects to produce robust and specific hybridization signals from RNA targets (8). Targets containing modified bases may have increased affinity towards a number of perfect match probes previously giving a poor hybridization signal. This would expand the variety of oligonucleotide probes which could be used in these experiments and allow increased freedom in selecting probes placed in strategic positions (i.e. splice junction sequences so as to monitor the expression of differentially spliced RNA transcripts).

Modified RNA transcripts can also be used to analyze the biophysical properties of nucleic acid structures (17,18,30). Others have incorporated modified bases into RNA during *in vitro* selection assays to expand the repertoire of RNA species having affinities for small molecules or specific catalytic properties (31). Increasing the stability of A·U base pairs using this strategy could expand the variety of nucleic acid structures which may have distinct biophysical properties.

ACKNOWLEDGEMENTS

We would like to thank Larry Brody and Bryan Sun for thoughtful discussion. These studies have been supported in part by research grant 5POLHGO1323-03 to SPAF.

REFERENCES

1 Fodor,S.P.A., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T. and Solas,D. (1991) *Science*, **251**, 767–773.
2 McGall,G., Labadie,J., Brock,P., Wallraff,G., Nguyen,T. and Hinsberg,W. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13555–13560.
3 Cronin,M.T., Fucini,R.V., Kim,S.M., Masino,R.S., Wespi,R.M. and Miyada,C.G. (1996) *Hum. Mutat.*, **7**, 244–255.
4 Kozal,M.J., Shah,N., Yang,R., Fucini,R., Merigan,T.C., Richman,D.D., Morris,D., Hubbell,E. and Gingeras,T.R. (1996) *Nature Med.*, **2**, 753–759.
5 Yershov,G., Barsky,R., Belgovskiy,A., Kirillov,E., Kreindlin, Ivanov,I., Parinov,S., Guschin,D., Drobishev,A., Dubiley,S. and Mirzabekov,A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 4913–4918.
6 Chee,M., Yang,R., Hubbell,E., Huang,X.C., Stern,D., Winkler,J., Lockhart,D.J., Morris,M.S. and Fodor,S.P. (1996) *Science*, **274**, 610–614.
7 Hacia,J.G., Brody,L.C., Chee,M.S., Fodor,S.P.A. and Collins,F.S. (1996) *Nature Genet.*, **14**, 441–447.
8 Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,K.M., Wang,C., Kobayashi,M., Horton,H. and Brown,E.L. (1996) *Nature Biotech.*, **14**, 1675–1680.
9 Shoemaker,D.D., Lashkari,D.A., Morris,D., Mittmann,M. and Davis,R.W. (1996) *Nature Genet.*, **14**, 450–456.
10 Milner,N., Mir,K.U. and Southern,E.M. (1997) *Nature Biotech.*, **15**, 537–541.
11 Hacia,J.G., Makalowski,W., Edgemon,K., Erdos,M.R., Robbins,C.M., Fodor,S.P.A., Brody,L.C. and Collins,F.S. (1998) *Nature Genet.*, **18**, 155-158.
12 Hobbs,F.W. (1989) *J. Org. Chem.*, **54**, 3420–3422.
13 Sowa,T. and Ouchi,S. (1975) *Bull. Chem. Soc. Jpn*, **48**, 2084–2090.
14 Ludwig,J. (1981) *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
15 Howard,F.B. and Miles,H.T. (1984) *Biochemistry*, **23**, 6723–6732.
16 Moffatt,J.G. and Khorana,H.G. (1961) *J. Am. Chem. Soc.*, **83**, 649–658.
17 Piccirilli,J.A., Krautch,T., Moroney,S.E. and Benner,S.A. (1990) *Nature*, **343**, 33–37.
18 Tor,Y. and Dervan,P.B. (1993) *J. Am. Chem. Soc.*, **115**, 4461–4467.
19 Zimmer,M. and Scheit,K.H. (1984) *Nucleic Acids Res.*, **12**, 2243–2258.
20 Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, NY.
21 Froehler,B.C., Wadwani,S., Terhorst,T.J. and Gerrard,S.R. (1992) *Tetrahedron Lett.*, **33**, 5307–5310.
22 Guitierrez,A.J., Matteucci,M.D., Grant,D., Matsumura,S., Wagner,R.W. and Froehler,B.C. (1997) *Biochemistry*, **36**, 743–748.
23 Chollet,A. and Kawashima,E. (1988) *Nucleic Acids Res.*, **16**, 305–317.
24 Chaejoon,C., Tinoco,I. and Chollet,A. (1988) *Nucleic Acids Res.*, **16**, 5115- 5122.
25 Melchior,W.B. and von Hippel,P.H. (1973) *Proc. Natl Acad. Sci. USA*, **70**, 298–302.
26 Rees,W.A., Yager,T.D., Korte,J. and von Hippel,P.H. (1993) *Biochemistry*, **32**, 137–144.
27 Ricelli,P.V. and Benight,A.S. (1993) *Nucleic Acids Res.*, **21**, 3785–3788.
28 Ratmeyer,L., Vianyak,R., Zhong,Y.Y., Zon,G. and Wilson,W.D. (1994) *Biochemistry*, **33**, 5298–5304.
29 Hoheisel,J.D. (1996) *Nucleic Acids Res.*, **24**, 430–432.
30 Moriyama,K., Negishi,K., Briggs,M.S.J., Smith,C.L., Hill,F., Churcher,M.J., Brown,D.M. and Loakes,D. (1998) *Nucleic Acids Res.*, **26**, 2105–2111.
31 Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) *Nature*, **389**, 54–57.